

The Proton Radius from Bayesian Inference

Krzysztof M. Graczyk* and Cezary Juszczak

Institute for Theoretical Physics, University of Wrocław, pl. M. Borna 9, 50-204, Wrocław, Poland

The methods of Bayesian statistics are used to extract the value of the proton radius from the elastic ep scattering data in a model independent way. To achieve that goal a large number of parametrizations (equivalent to neural network schemes) are considered and ranked by their conditional probability $P(\text{parametrization} | \text{data})$ instead of using the minimal error criterion. As a result the most probable proton radii values ($r_E^p = 0.899 \pm 0.003$ fm, $r_M^p = 0.879 \pm 0.007$ fm) are obtained and systematic error due to freedom in the choice of parametrization is estimated. Correcting the data for the two photon exchange effect leads to smaller difference between the extracted values of r_E^p and r_M^p . The results disagree with recent muonic atom measurements.

PACS numbers: 13.40.Gp, 25.30.Bf, 14.20.Dh, 84.35.+i

Keywords: proton radius, proton form-factors, two-photon exchange correction

The problem of the proton radius is being discussed in the particle and atomic physics for many years but recently it has received even more attention due to the new results coming from the Lamb shift measurements of the muonic hydrogen atom (μp) [1]. It is a very accurate estimate of the proton radius but it is inconsistent with the CODATA value (compilation of the measurements in hydrogen atom and analysis of the electron-proton scattering data) [2]. This disagreement, called “the proton radius puzzle”, has not been explained yet. There are many proposals to resolve the problem including considering different types of interaction for ep and μp [3] or “weakening the assumptions of perturbative formulation of quantum electrodynamics within the proton” [4]. In this paper we introduce a novel approach to the extraction of the proton radius from the ep scattering data, which allows one to control the model dependence of the result within a Bayesian objective algorithm.

The electromagnetic (E-M) structure of the proton is encoded in the electric G_E and magnetic G_M form factors, which are also the crucial input for the atomic (hydrogen-like) calculations [5]. In the low Q^2 limit in the Breit frame they are related with the distributions of the electric charge and magnetic momentum inside the nucleon [6]. In particular the nucleon radius is expressed by the slope of the form factors at vanishing Q^2 :

$$r_{E,M}^p \equiv \left(-\frac{6}{G_{E,M}(0)} \frac{dG_{E,M}(Q^2)}{dQ^2} \Big|_{Q^2=0} \right)^{\frac{1}{2}}. \quad (1)$$

The value of $r_{E,M}^p$ can be obtained from the scattering data (mainly elastic ep). Recent results include [7–11] and a more complete list can be found in Refs. [3, 12, 13]. These analyses have been performed in the spirit of the frequentistic statistics.

The values of the proton radius (related to the electric charge distribution) obtained by different authors during

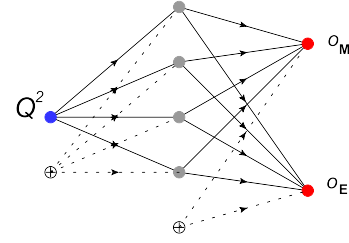


FIG. 1: (Color online) The scheme of the neural network (with $H = 4$ hidden units) used in the analysis, for reference see Eq. 6. Every line (connection) corresponds to one parameter w_i . Gray circles denote the sigmoid functions, open circles are constants.

the last fifty years, range from about 0.8 fm to about 0.9 fm [14]. Indeed, as it was demonstrated in [8, 9] the results depend the choice of the form factor parametrization but it is obvious that the collection of datasets used in the analysis also matters. Even the right choice of the number of parameters in a given class of parametrizations can be a challenge. More parameters in the fit can always reduce the χ^2 error but at some point the fit tends to reproduce statical fluctuations of the experimental measurements. Such a model overfits the data i.e. describes existing points with unrealistic precision but has no predictive power – introducing new data points drastically increases the χ^2 error and spoils the quality of the fit. This is connected with the so called bias variance trade-off. Some attempt to resolve this problem has been made in Ref. [9] where conformal mapping was used in order to exploit the analytic properties of the form factors.

Our philosophy is different. In contrast to the frequentistic statistics methods used in the above papers, we use the Bayesian statistics which is well suited to model comparison. Within this framework, adapted for neural networks, we select the most optimal model (according to the data) and extract the value of the proton radius.

The extraction of the nucleon form factors in [15] was our first use of the Bayesian neural network framework. In the next paper [16] the approach was significantly im-

*Electronic address: krzysztof.graczyk@ift.uni.wroc.pl

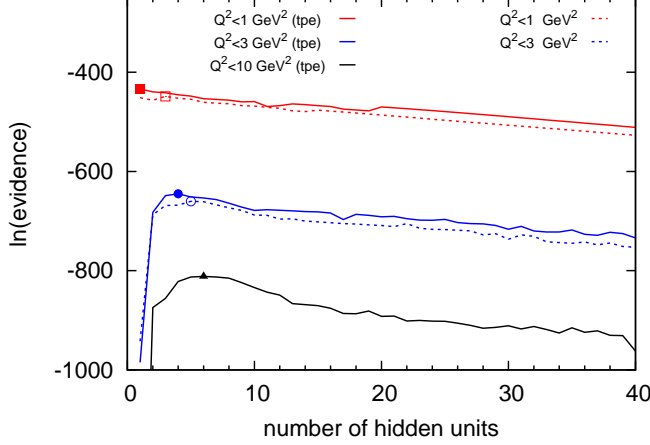


FIG. 2: (Color online) Logarithm of evidence for the best neural networks in each scheme. Solid/dashed lines correspond to the analyses of the data corrected/not corrected by the TPE effect. Points mark the best models for each analysis.

proved to obtain the two-photon exchange (TPE) correction to unpolarized elastic ep cross section. However, because of some physical assumptions these analyses were not dedicated to the low- Q^2 region. In the present paper we overcome this difficulty because: (i) the TPE contribution (elastic and inelastic described by box diagrams with nucleon and $\Delta(1232)$ resonance as intermediate hadronic states) calculated in [17] is subtracted from the data and only E-M form factors are extracted from the cross section and polarization transfer data; (ii) the form factors are re-parameterized in order to automatically fulfil $G_E(0) = G_M(0)/\mu_p = 1$ (μ_p is the proton magnetic moment); (iii) our software has been significantly improved in terms of accuracy and efficiency.

Here we briefly outline the Bayesian framework, for more details see [15–17]. The main idea is to consider a large class of models. By a model we mean the function \mathcal{N} used to fit the data \mathcal{D} and two conditional probabilities: $P(\{w_i\}|\mathcal{N})$ – prior which accommodates the initial assumption about the $\{w_i\}$, and the likelihood $P(\mathcal{D}|\{w_i\}, \mathcal{N})$; $\{w_i\}$ is the set of parameters of the function \mathcal{N} . Then the models are ranked by the conditional probability $P(\mathcal{N}|\mathcal{D})$, which can be replaced by the evidence $P(\mathcal{D}|\mathcal{N})$ if the data is the same and no model is preferred at the beginning of the analysis.

For any given model we denote with $\{w_i\}_{MP}$ the configuration of the parameters which maximizes the posterior

$$P(\{w_i\}|\mathcal{D}, \mathcal{N}) = \frac{P(\mathcal{D}|\{w_i\}, \mathcal{N})P(\{w_i\}|\mathcal{N})}{P(\mathcal{D}|\mathcal{N})}. \quad (2)$$

Finding $\{w_i\}_{MP}$ in one of the steps of the analysis.

Usually the evidence is peaked at $\{w_i\}_{MP}$ and it simplifies to the likelihood at the maximum multiplied by the Occam factor which makes too complex models less

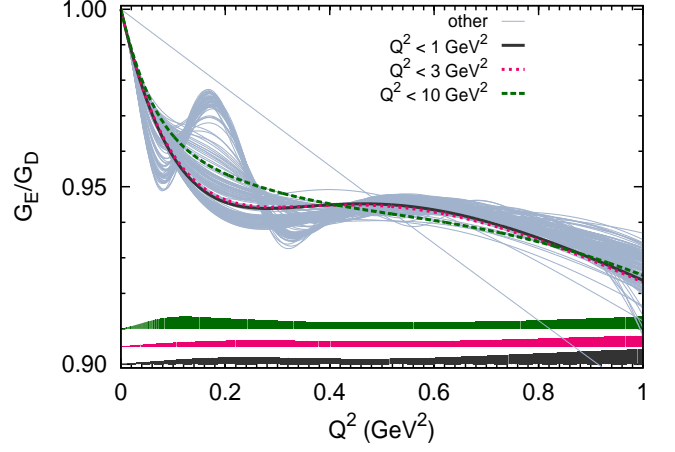


FIG. 3: (Color online) The form factor G_E/G_D ($G_D = 1/(1 + Q^2/0.71\text{GeV}^2)^2$). Red, blue and magenta lines correspond to the best models in the analysis with data limited by $Q^2_{\text{cutoff}} = 1, 3$ and 10 GeV^2 respectively. The areas colored with magenta, blue and red denote 1σ uncertainty due to the change in the fit parameters (calculated within the Hessian approximation). The gray lines show models which are best for neural networks with definite number of hidden units, but are not globally best.

likely [18]

$$P(\mathcal{D}|\mathcal{N}) \approx P(\mathcal{D}|\{w_i\}_{MP}, \mathcal{N}) \underbrace{(2\pi)^{\frac{W}{2}} (\det A)^{-\frac{1}{2}}}_{\text{Occam factor}}, \quad (3)$$

where W is the number of parameters $\{w_i\}$ and $A_{ij} = -\nabla_{w_i} \nabla_{w_j} \ln P(\{w_i\}|\mathcal{D}, \mathcal{N})|_{\{w_i\}=\{w_i\}_{MP}}$.

For the function \mathcal{N} we take the feed-forward neural networks with one hidden layer of units (Fig. 1), which correspond to linear combinations of sigmoid functions. This class of functions can be used to approximate any continuous function with arbitrary precision [19].

The prior function for the neural network parameters $\{w_i\}$ has the standard form $\mathcal{P}(\{w_i\}|\alpha, \mathcal{N}) \sim e^{-\alpha E_w}$ where $E_w = \frac{1}{2} \sum_{i=1}^W w_i^2$. The parameter α is the regularizer which determines the width of the initial Gaussian distribution for the parameters. In principle it should be treated as another parameter of the model and its optimal value α_{MP} is also obtained during the analysis. The likelihood is defined by the χ^2 distribution

$$\mathcal{P}(\mathcal{D}|\{w_i\}, \alpha, \mathcal{N}) \sim e^{-\chi_{cr.}^2(\mathcal{D}, \{w_i\}) - \chi_{PT}^2(\mathcal{D}, \{w_i\})}. \quad (4)$$

The $\chi_{cr.}^2$ includes the unpolarized elastic ep scattering cross section data, which are assumed to be parametrized by one-photon exchange formula $\sigma_R = Q^2 G_M^2 / 4M_p^2 + \varepsilon G_E^2$ (M_p is the proton mass, ε denotes photon polarization). Each independent set of cross section measurements is characterized by its systematic normalization uncertainty so we introduce a separate normalization parameter for every data set. Values of these parameters are

obtained by an iterative algorithm described in [16]. The χ_{PT}^2 includes the form factor ratio $\mu_p G_E/G_M$ data from the polarization transfer measurements. We consider the same selection of data as in [16] with one additional data set from [20] and make the prior assumption that all data sets are equally relevant. The MAMI data [7] are not included in the analysis because of the difficulties described in [21].

The optimal configuration $\{\alpha, w_i\}_{MP}$ is found by an iterative procedure based on the fact that the maximum of the posterior corresponds to the minimum over $\{w_i\}$ of the error function $S(\mathcal{D}, \{w_i\}) \equiv \chi_{cr.}^2(\mathcal{D}, \{w_i\}) + \chi_{PT}^2(\mathcal{D}, \{w_i\}) + \alpha_{MP} E_w$ and also $\frac{\partial}{\partial \alpha} \mathcal{P}(\mathcal{D} | \alpha, \mathcal{N})|_{\alpha=\alpha_{MP}} = 0$, where

$$\mathcal{P}(\mathcal{D} | \alpha, \mathcal{N}) = \int dw \mathcal{P}(\mathcal{D} | \{w_i\}, \alpha, \mathcal{N}) \mathcal{P}(\{w_i\} | \alpha, \mathcal{N}).$$

Eventually the logarithm of evidence for the model \mathcal{N} (in the Hessian approximation) reads:

$$\ln P(\mathcal{D} | \mathcal{N}) \approx -S(\mathcal{D}, \{w_i\}_{MP}) + \frac{W}{2} \ln \alpha_{MP} - \frac{1}{2} \ln \det A - \frac{1}{2} \ln \sum_{i=1}^W \frac{1}{2} \frac{\lambda_i}{\lambda_i + \alpha_{MP}} \quad (5)$$

where $A_{kj} = \nabla_{w_k} \nabla_{w_j} S(\mathcal{D}, \{w_i\})$ and λ_k -s are the eigenvalues of the matrix $\nabla_{w_i} \nabla_{w_j} (\chi_{cross}(\mathcal{D}, \{w_i\}) + \chi_{PT}(\mathcal{D}, \{w_i\}))$.

As the form factors G_E , G_M describe the properties of the same object, they must be related, so we parametrize them with a single feed forward neural network with one input (Q^2) and two outputs, o_E and o_M (see Fig. 1),

$$G_E = (1 - Q^2 o_E) G_D, \quad G_M = \mu_p (1 - Q^2 o_M) G_D, \quad (6)$$

where $G_D(Q^2) = (1 + Q^2/M_V^2)^{-2}$ and $M_V^2 = 0.71 \text{ GeV}^2$. Then the value of the proton radius is given by

$$\left(r_{E,M}^p\right)^2 = \left(r_{dipole}^p\right)^2 + 6 o_{E,M}(Q^2 = 0), \quad (7)$$

where $(r_{dipole}^p)^2 = 12/M_V^2$. We see that the network outputs at $Q^2 = 0$ directly contribute to the deviation of the proton radius from the dipole form factor value.

In Refs. [14, 22] it was shown that modifying the cross section data by the Coulomb distortion (CD) is important in the extraction of the proton radius. On the other hand, the CD is a part of the TPE effect¹, which has to be subtracted from the data to obtain the form factors which agree with the polarization transfer measurements

| $Q_{cutoff}^2 \text{ (GeV}^2\text{)}$ | $r_M^p \text{ (fm)}$ | $r_E^p \text{ (fm)}$ | H |
|---------------------------------------|----------------------|----------------------|-----|
| 1 | 0.879 ± 0.007 | 0.899 ± 0.003 | 1 |
| 3 | 0.883 ± 0.007 | 0.899 ± 0.003 | 4 |
| 10 | 0.953 ± 0.065 | 0.897 ± 0.005 | 6 |

TABLE I: The values of the proton radii obtained with 1σ uncertainty due to variation in the parameter space. H is the number of hidden units of the best model.

[26]. Similarly as in [27] we correct the cross section data by the TPE contribution (calculated in [17] in similar way as in [28, 29]).

The set of all possible parametrizations is infinite. In our approach the models are grouped according to the number of units in the hidden layer. For each group we found the $\{w_i, \alpha\}$ which maximize the evidence. The dependence of the maximal evidence on the number of hidden units is plotted in Fig. 2. For every analysis the evidence reaches a peak (which defines the best model) and then starts to fall. Because of this fact we do not consider the networks with more than 40 hidden units. The total number of obtained neural network parameterizations exceeded half million.

It needs to be underlined that the models which maximize the evidence are not those which minimize the χ^2 . This is true in each class of models and also globally. The χ_{min}^2 decreases with the number of parameters possibly saturating at some point. But the models with the lowest χ^2 tend to overfit the data, which is one of the reasons why the χ_{min}^2 is not a suitable criterion for ranking the models. The other problems of χ^2 -based criteria are discussed in [3]. Objective mathematical methods for model comparison are provided by the Bayesian statistics.

In Fig. 3 it is shown how the form factor plots depend on the choice of the parametrization. The gray lines correspond to the best models in each group, while the globally best for each analysis are shown with color. Certainly the choice of the form factor parametrization has also strong impact on the obtained proton radius value (Fig. 4). Hence it is crucial to have criterion for finding the model which is the most favorable by the measurements. On the other hand each model can be true with the probability $P(\mathcal{N} | \mathcal{D})$ (proportional to the evidence) so it is possible to calculate the expected value of the proton radius and the systematic uncertainty due to the choice of the model. In practice the expected value is very close to the most probable value (see Tables I and II).

In Fig. 4 we show the proton radii values obtained from the analyses where data points with Q^2 above $Q_{cutoff}^2 = 1, 3$ and 10 GeV^2 are rejected (for the data uncorrected by the TPE we considered only $Q_{cutoff}^2 = 1, 3 \text{ GeV}^2$). The choice of the value for the cutoff has small impact on the extraction of the proton radii. For the result of our analysis we take those from the lowest cut, namely $r_E^p = 0.899 \pm 0.003 \text{ fm}$ and $r_M^p = 0.879 \pm 0.007 \text{ fm}$.

¹ The relevance of the TPE effect and higher order Born corrections in the proton radius extraction is also discussed in Refs. [23] and [24, 25] respectively.

| Q_{cutoff}^2 (GeV ²) | r_M^p (fm) | r_E^p (fm) |
|---|---------------------|---------------------|
| 1 | 0.8792 ± 0.0006 | 0.8989 ± 0.0001 |
| 3 | 0.8828 ± 0.0063 | 0.8988 ± 0.0003 |
| 10 | 0.9205 ± 0.0606 | 0.8968 ± 0.0029 |

TABLE II: The expected value (according to evidence probability distribution) of the proton radii with systematic uncertainty due to the choice of the parametrization (given by the variance).

On the other hand the r_E^p 's obtained based on the minimal error criterion depend on the cutoff choice. These results are smaller than those indicated by the evidence and have larger uncertainties, see Fig. 4. We leave for future studies within the Bayesian framework the quantitative investigation of the dependence of the results on the data set selection and Q_{cutoff}^2 value. Eventually it can be seen in Fig. 4 that using the data uncorrected by TPE effect leads to larger uncertainties of r_E^p and r_M^p and larger difference $r_E^p - r_M^p$ (0.02 fm with TPE and 0.09 fm without).

We would like to emphasize that according to our knowledge the present work is the first extraction of the proton radius based on the Bayesian methods. The extracted value of the proton radius r_E^p agrees with previous results [9, 14, 27] but disagrees with the muonic atom measurements [1]. The subtracting of the TPE correction from the cross section data plays an important role in the analysis.

The calculations have been carried out in Wrocław Centre for Networking and Supercomputing (<http://www.wcss.wroc.pl>), grant No. 268.

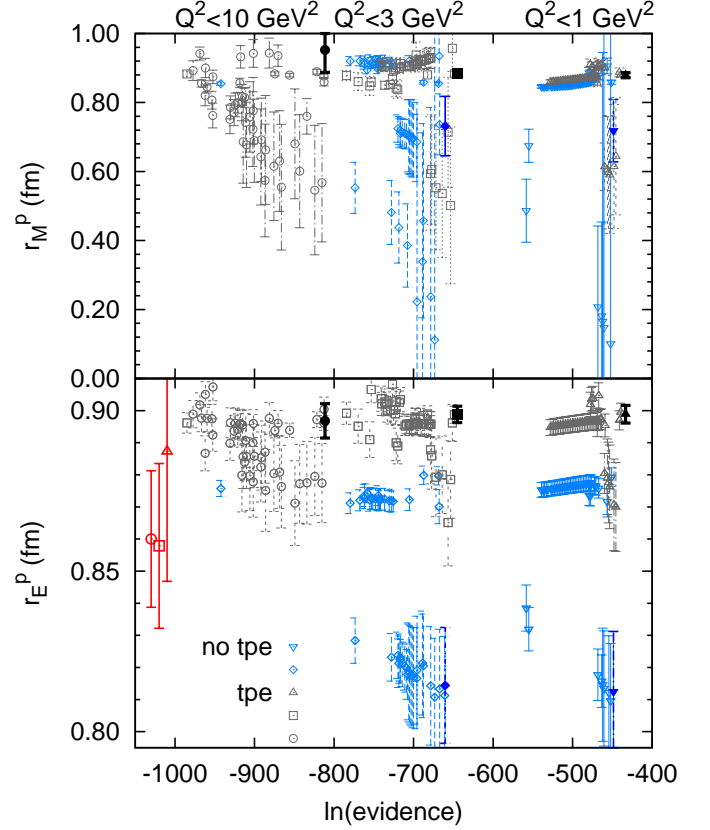


FIG. 4: (Color online) The proton radii corresponding to the models which are the best within particular data selection and neural network scheme. The results for the data corrected by the TPE are marked with black \triangle , \square , and \circ for $Q_{\text{cutoff}}^2 = 1, 3$ and 10 GeV² respectively. The results for the uncorrected data are marked with blue ∇ and \diamond for $Q_{\text{cutoff}}^2 = 1$ and 3 GeV² respectively. The filled points mark the best result for each case. The three leftmost red points are the best results according to the minimum of the error function. For clarity their x-coordinate has been changed. Their $\ln(\text{evidence})$ values are -718 , -1070 , and -1311 . The picture is based on the analysis of $39 \cdot 5 = 195$ models.

[1] R. Pohl, A. Antognini, F. Nez, F. D. Amaro, F. Biraben, *et al.*, *Nature* **466**, 213 (2010).
[2] P. J. Mohr, B. N. Taylor, and D. B. Newell, *Rev.Mod.Phys.* **84**, 1527 (2012), arXiv:1203.5425 [physics.atom-ph].
[3] E. Kraus, K. Mesick, A. White, R. Gilman, and S. Strauch, (2014), arXiv:1405.4735 [nucl-ex].
[4] K. Pachucki and K. A. Meissner, (2014), arXiv:1405.6582 [hep-ph].
[5] K. Pachucki, *Phys.Rev.* **A53**, 2092 (1996).
[6] F. Ernst, R. Sachs, and K. Wali, *Phys.Rev.* **119**, 1105 (1960).
[7] J. Bernauer *et al.* (A1 Collaboration), *Phys.Rev.Lett.* **105**, 242001 (2010), arXiv:1007.5076 [nucl-ex].
[8] J. Bernauer *et al.* (A1 Collaboration), (2013), 10.1103/PhysRevC.90.015206, arXiv:1307.6227 [nucl-ex].
[9] R. J. Hill and G. Paz, *Phys.Rev.* **D82**, 113005 (2010), arXiv:1008.4619 [hep-ph].
[10] I. Lorenz, H.-W. Hammer, and U.-G. Meissner, *Eur.Phys.J.* **A48**, 151 (2012), arXiv:1205.6628 [hep-ph].

[11] I. Lorenz and U. G. Meissner, (2014), arXiv:1406.2962 [hep-ph].
[12] Z. Epstein, G. Paz, and J. Roy, (2014), arXiv:1407.5683 [hep-ph].
[13] S. G. Karshenboim, (2014), arXiv:1405.6515 [hep-ph].
[14] I. Sick, *Phys.Lett.* **B576**, 62 (2003), arXiv:nucl-ex/0310008 [nucl-ex].
[15] K. M. Graczyk, P. Plonski, and R. Sulej, *JHEP* **1009**, 053 (2010), arXiv:1006.0342 [hep-ph].
[16] K. M. Graczyk, *Phys.Rev.* **C84**, 034314 (2011), arXiv:1106.1204 [hep-ph].
[17] K. M. Graczyk, *Phys.Rev.* **C88**, 065205 (2013), arXiv:1306.5991 [hep-ph].
[18] D. MacKay, *Bayesian Methods for Adaptive Models*, Ph.D. thesis, California Institute of Technology (1991).
[19] G. Cybenko, *Math Control, Signal* **2**, 303 (1989).

- [20] X. Zhan, K. Allada, D. Armstrong, J. Arrington, W. Bertozzi, *et al.*, Phys.Lett. **B705**, 59 (2011), arXiv:1102.0318 [nucl-ex] .
- [21] J. Arrington, Phys.Rev.Lett. **107**, 119101 (2011), arXiv:1108.3058 [nucl-ex] .
- [22] R. Rosenfelder, Phys.Lett. **B479**, 381 (2000), arXiv:nucl-th/9912031 [nucl-th] .
- [23] M. Gorchtein, (2014), arXiv:1406.1612 [nucl-th] .
- [24] J. Arrington, J.Phys. **G40**, 115003 (2013), arXiv:1210.2677 [nucl-ex] .
- [25] J. Arrington and I. Sick, Phys.Rev. **C70**, 028203 (2004), arXiv:nucl-ex/0406014 [nucl-ex] .
- [26] J. Arrington, W. Melnitchouk, and J. Tjon, Phys.Rev. **C76**, 035205 (2007), arXiv:0707.1861 [nucl-ex] .
- [27] P. G. Blunden and I. Sick, Phys.Rev. **C72**, 057601 (2005), arXiv:nucl-th/0508037 [nucl-th] .
- [28] P. Blunden, W. Melnitchouk, and J. Tjon, Phys.Rev. **C72**, 034612 (2005), arXiv:nucl-th/0506039 [nucl-th] .
- [29] S. Kondratyuk, P. Blunden, W. Melnitchouk, and J. Tjon, Phys.Rev.Lett. **95**, 172503 (2005), arXiv:nucl-th/0506026 [nucl-th] .